# LETTERS TO THE EDITOR

## RESTRICTED RANDOM MATING: A NEW MATING MODEL

ESTIMATION of gene frequencies for any genetic character involves always directly or indirectly some assumptions about the structure of the population concerned (It may be noted, however, that for characters where all genotypes are phenotypically distinguishable, estimation of gene frequencies by gene count does not involve any assumption about the genetic structure of the population.) Furthermore, we know that construction of a model for studying the structure of a population again assumes the prevailing mating scheme in the population.[1] Kempthorne[2] has shown that repeated random mating leads to a population prescribed by Model I (Li and Horvitz[3]) and when a random mating population is inbred to an extent of F ($0 \leqslant F \leqslant 1$), the resulting population behaves as a Model II (Wright's equilibrium population) one. Though these two are the most frequently studied population structures, their applicability is very often questioned.[4] Here, in this note, our main contention is to announce that one can proceed without any assumption regarding the mating structure at the phenotypic level and later use such a set-up to compute gene frequencies from a two-generation data.

To start with, let us consider two alleles 'A' and '$a$' at an autosomal locus. Let us also assume that 'A' is dominant over '$a$' so that only two phenotypes $\bar{A}$ (representing genotypes AA and $Aa$ collectively) and $\bar{a}$ (representing the genotype $aa$) are distinguishable. Thus, in this system, there are three phenotypic mating types. We designate the probabilities of the mating types $\bar{A} \times \bar{A}$, $\bar{A} \times \bar{a}$ and $\bar{a} \times \bar{a}$, by $\lambda_1$, $\lambda_2$ and $\lambda_3$ respectively ($0 \leqslant \lambda_i \leqslant 1$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$). Note that the attempt of writing these $\lambda_i$'s as functions of $p$ and $q$ (the frequencies of the gene A and $a$, respectively) needs assumption about the mating structure. Instead, we define a parameter $\theta$ as the conditional probability that an individual is of genotype $Aa$ given that his phenotype is known to be $\bar{A}$.

With these parameters at hand and the general set-up at the phenotypic level, we make use of Hardy-Weinberg Law only at the level of dividing the general phenotypic mating frequencies into the corresponding genotypic frequencies. As for example, consider the mating type $\bar{A} \times \bar{A}$. This phenotypic mating can be split into the three corresponding genotypic matings in the proportions as shown in Table I.

TABLE I

*Genotypic mating types given the phenotypic mating $\bar{A} \times \bar{A}$ and their probabilities*

| Genotype mating type | Probability |
|---|---|
| $Aa \times Aa$ | $\theta^2$ |
| $Aa \times AA$ | $2\theta\,(1-\theta)$ |
| $AA \times AA$ | $(1-\theta)^2$ |

With these it is easy to see that the segregation probabilities can be designated by Table II as follows (the N's and $n$'s are the corresponding observed frequencies in a random sample of N families).

TABLE II

*Parental matings and the offspring phenotype probabilities*

| Parental matings | | Offspring | |
|---|---|---|---|
| Type | Frequency | $\bar{A}$ | $\bar{a}$ |
| $\bar{A} \times \bar{A}$ | $\lambda_1$ ($N_1$) | $1-\alpha^2$ ($n_{11}$) | $\alpha^2$ ($n_{12}$) |
| $\bar{A} \times \bar{a}$ | $\lambda_2$ ($N_2$) | $1-\alpha$ ($n_{21}$) | $\alpha$ ($n_{22}$) |
| $\bar{a} \times \bar{a}$ | $\lambda_3$ ($N_3$) | 0 | 1 ($n_{32}$) |
| Total | N | | |

where $\alpha = \theta/2$.

Once this is done, the estimation is carried out by maximum likelihood method and it may be observed that the estimates of $\lambda_i$'s are nothing but the observed relative frequencies ($N_i/N$) of the three phenotypic mating types. Estimate of $\theta$ depends upon the observations on the children only ($\hat{\theta} = 2\hat{\alpha}$, where $\hat{\alpha}$ is the positive root of $c_1\alpha^2 + c_2\alpha - c_3 = 0$; $c_1 = 2n_{11} + 2n_{12} + n_{21} + n_{22}$, $c_2 = n_{21}$ and $c_3 = 2n_{12} + n_{22}$) and they are asymptotically uncorrelated with the estimates of $\lambda_i$'s. To test the goodness of fit statistic is evolved and its

large sample distribution will be presented elsewhere.[5]

It is to be remembered that once the $\theta$ value is known for a population, the estimation of $p$ and $q$ can be carried out by means of gene counts, since the expected genotypic proportions will be known (*e.g.*, the relative frequency of $Aa$ genotypes $= \theta \times$ relative frequency of phenotype $\bar{A}$).

In this general analysis, it is to be noted that, the hypothesis of random mating is not dispensed with altogether. However, since in most cases there is no easy way of detecting an $\bar{A}$ individual as $AA$ or $Aa$ (in case of blood groups, specially) any model which takes this point into account is going to be only of mathematical interest. So, the assumption of random mating in a restricted sense (at genotypic level only) as outlined above does not reduce the generality of the model.

Indian Statistical Inst., RANAJIT CHAKRABORTY.
203, B.T. Road,
Calcutta-35, *February* 9, 1971.

1. Chakraborty, R., Unpublished Ph.D. Thesis at Indian Statistical Institute, Calcutta, 1970.
2. Kempthorne, O., *Introduction to Genetic Statistics*, John Wiley and Sons, 1957 edition, Chapter II.
3. Li, C. C. and Horvitz, D. G., *Amer. J. Hum. Genet.*, 1953, **5**, 107.
4. Smith, C. A. B., Personal communication, 1970.
5. Adhikari, B. P., Sarma, Y. R. and Chakraborty, R., Unpublished, 1970.